UNCLASSIFIED

(4)

CSB Working Paper RESEARCH 11
CENTRAL STUDIES BRANCH
Author(s): B.K. McMILLAN

JULY 1988
AR-005-268

COPY NO. 43 OF 33

CHI-SQUARED DISTRIBUTION PARAMETER APPROXIMATION

To all Recipients

1. This paper has been prepared as a contribution to the ADSES Simulation.

2. This working paper is a first issue for comment, oral or written, which should be directed to B.K. McMillan (66 4311) by 31 August 1988.

3. CSB working papers are intended as a basis for discussion on some specific aspect of a study, or to inform recipients of intermediate or final results of a study. They are not to be referenced in offical documents without referral to CSB. Working papers which are no longer required by recipients should be destroyed in accordance with Public Service security regulations.

UNCLASSIFIED

CONDITIONS OF RELEASE AND DISPOSAL

1. This document is the property of the Australian Government. The information it contains is released for defence purposes only and must not be disseminated beyond the stated distribution without prior approval.

2. The document and the information it contains must be handled in accordance with security regulations applying in the country of lodgement, downgrading instructions must be observed and removal of any of the above limitations is only with the specific approval of the Releasing Authority as given in the Secondary Distribution statement.

3. This information may be subject to privately owned rights.

4. The officer in possession of this document is responsible for its safe custody. When no longer required this document should be destroyed in accordance with Public Service security regulations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

1. This paper gives a method for approximating the mean of a chi-squared (χ^2) distribution as well as an offset constant and a multiplicative constant to be applied to variates from that distribution. They must be estimated from the given values of a median, a lower decile and an upper decile.

2. The origin of the problem is in OTHR modelling, where day to day variations for the 'best' frequency are approximated by a statistical distribution, after explainable elements have been removed. In addition the distribution reverses its skewness under certain circumstances. The model accuracy is uncertain, and variability in observed results is high. Consequently accurate parameter assessment is seen as unnecessary.

3. Data for parameter estimation come from the OTHR modelling process and long term observations. That is, the physics involved have been used to produce a model describing the way in which the median maximum useable frequency (MUF) should vary. Observations of the actual MUF on a day to day basis have suggested that the χ^2 distribution suitably transformed describes actual frequency variations about the frequencies predicted by the model. Figure 1 illustrates how the median might vary on a day to day basis, and how the actual frequencies might vary about this line.

4. The lower bound of the χ^2 distribution is zero, and its mean describes the extent to which it is skewed. To give a better fit to the data it is possible to multiply by a constant, thus preserving the skewness, but allowing variation in the mean. Another factor that can be introduced to improve the fit is an additive constant. This allows the data to tail off at some value other than 0. Both of these elements have been introduced in the modelling process.

5. As stated above, the physics modelling process produces a median MUF (rather than a mean). From this figure, estimates of the highest and lowest deciles are produced simply by applying multiplicative constants to the median. The constants have values that depend on time of day, year and sunspot cycle.

NOTATIONS AND DEFINITIONS

F	Frequency.
F_i	Frequency deciles. In particular i=1, 5, 9.
$X(M)$	Variate from a χ^2 distribution mean M.
x_i	Decile values of X(M).
e_i	Constants associated with x_i.
c,d	Constants.
A,B	Constants.

METHODOLOGY AND RESULTS

6. In accordance with Reference 1 the following relationship is assumed:

$$F = A + B.X(M) \tag{1}$$

7. The problem is to estimate A, B, and M from F_1, F_5, and F_9 with a minimum of effort and at a level of accuracy which is unspecified but not necessarily high. As far as practical, the varying degree of skewness of the distribution with M should not affect the result. In addition it is expected that most of the time M should be more than 5.

8. Examination of tables for the χ^2 distribution yielded the following approximate relationships:

$$x_5 = M + e_5 \tag{2}$$

$$x_1 + x_9 = 2.M + e_9 \tag{3}$$

$$M = c.(x_1 - x_9)^d \tag{4}$$

Choosing different values for e_i, c and d according to M over or under 5 improved the approximation. Table 1 gives details.

9. From these relationships it is possible to obtain values for A, B, and M using F_i and Equations 1 to 4:

$$F_1 - F_9 = B(x_1 - x_9)$$

$$= B(M/C)$$

$$F_1 + F_9 - 2.F_5 = B.(x_1 + x_9 - 2.x_5)$$

$$= B.(e_9 - 2.e_5)$$

Whence $$B = \frac{F_1 + F_9 - 2.F_5}{e_9 - 2.e_5} \tag{5}$$

Thus $$M = c.\left(\frac{F_1 - F_9}{B}\right)^d \tag{6}$$

and $$A = F_5 - B.(M + e_5) \tag{7}$$

In the event that B is negative, the χ^2 distribution can be assumed to be skewed in the reverse direction. To deal with this, F_1 and F_9 may be interchanged in the calculations for A, B and M. The change only affects equation 6 where it cancels the negative sign of B, allowing M to have a real positive solution. In application, using equation 1, the value of B remains negative, so the scaled variates are subtracted from A, thus in effect reversing the distribution.

ACCURACY CONSIDERATIONS

10. Equations 2 to 4 are approximate relationships, and have an error caused by assuming e_i, c and d to be constants. These errors can be calculated for selected values of M using the constants values (Table 1) and the χ^2 deciles x_i of Table 2. Figure 2 shows the values chosen for e_5 and e_9 (solid lines) and the values they would take if they were not constants (dashed lines). The error is the difference between them, and expressed as a percentage of M the figure shows that for M over 5 they are less than about 2 per cent and mostly less than 0.2 per cent. For M under 5, the error can be as high as 7½ per cent for e_5 and almost 11 per cent for e_9 at $M = 1$, although such low values of M are expected to occur rarely.

11. Figure 3 shows the elements M of Equation 4 plotted on log by log paper. This paper was chosen to enable constants c and d to relate to the intercept and gradient respectively of a straight line. For given values of M, the actual values of $x_1 - x_9$ are plotted (x points) and the estimate of these values calculated using the inverse of Equation 4. The scale of this graph is such that it is harder to see the magnitude of the errors than for Figure 2, but it shows that for M greater than 5 the error as a proportion of M is less than 5 per cent and mostly less than 1 per cent. For M under 5 it reaches about 7½ per cent at $M = 2$ and 10 per cent at $M = 1$.

12. The effect of these errors is as follows:

a. In assessing B (Equation 5), the error is that of $e_9 - 2.e_5$ which as a proportion of M is 4 per cent at M = 1 and less than 1 per cent for M not less than 2.

b. In assessing M (Equation 6), the error is that of B to the power d, combined with those observed in Figure 2. Near M = 1 this could produce an error of 16 per cent of M but for most other cases it should be no more than 2 per cent and often will be much less.

c. Errors in A need to be considered in conjunction with the whole process, since its value is selected to force the approximation through F_5 (Equation 6). Thus in the centre of the range where most results can be expected, the approximation will be at its most accurate.

13. There are too many variable elements for a graphical representation of all errors. However a selected range of values of F_1, F_5 and F_9 have been used to estimate A, B and M using the above processes. These values have then been used to estimate the original F_1 and F_9 values. Because the largest errors occur when M is under 5, most emphasis has been placed on this region. Table 3 shows the results and demonstrates errors of under 5 per cent of F_5 at the extreme deciles. The columns headed F_1 and F_9 should be compared for this result. Columns x_1 and x_9 giving the χ^2 deciles were interpolated or extrapolated from table 2 for the values of M calculated. Equation 1 was used to calculate the estimates for F_1 and F_9.

SUMMARY

14. A quick and effective method for determining from three frequency deciles the mean of a χ^2 distribution and an associated linear transform has been presented. The original purpose of the calculations was to allow creation of suitably distributed OTHR frequencies to be obtained, but the general principles used are applicable in other areas either in whole or in part. Errors have been estimated and have been shown to be greatest for small values of the mean, e.g. possibly as high as 20 per cent of the mean, although in practice errors of less than 10 per cent for the central 80 per cent of the distribution can be expected. For most common values encountered in OTHR modelling, errors of less than 2 per cent of the mean can be expected.

REFERENCES

1. CCIR Report 252-2. Equation 20b.

2. Burington and May. Handbook of Probability and Statistics with Tables.

Table 1. VALUES OF FITTING CONSTANTS

CONSTANT	VALUE	
	$M \leq 5$	$M \geq 5$
e_5	-0.63	-0.66
e_9	0.83	0.85
c	0.21	0.106
d	1.55	.893

Table 2. χ^2 DECILES 1, 5 AND 9, AND COMBINATIONS

M	x_1	x_5	x_9	x_1+x_9	x_1-x_9	$x_1+x_9-2x_5$
1	0.016	0.455	2.706	2.722	2.690	1.812
2	0.211	1.386	4.605	4.816	4.394	2.044
3	0.584	2.366	6.251	6.835	5.667	2.103
4	1.064	3.357	7.779	8.843	6.715	2.129
5	1.610	4.351	9.236	10.846	7.626	2.144
6	2.204	5.348	10.645	12.849	8.441	2.153
7	2.833	6.346	12.017	14.850	9.284	2.158
8	3.490	7.344	13.362	16.852	9.872	2.164
9	4.168	8.343	14.648	18.852	10.516	2.166
10	4.865	9.342	15.987	20.852	11.122	2.168
12	6.304	11.340	18.549	24.853	12.245	2.172
15	8.547	14.339	22.307	30.854	13.760	2.176
20	12.443	19.337	28.412	40.855	15.969	2.181
25	16.473	24.337	34.382	50.855	17.909	2.181
30	20.599	29.336	40.256	60.855	19.657	2.183

Table 3. RESULTS AND ERRORS FOR SELECTED FREQUENCIES

SELECTED FREQUENCIES			CALCULATED PARAMETERS			X^2 DECILES FOR M		CALCULATED FREQUENCY DECILES	
F_1	F_5	F_9	B	M	A	x_1	x_9	F_1	F_9
1	2	10	3.35	0.97	0.861	0.015	2.62	0.91	9.6
1	3	10	2.39	1.64	0.586	0.11	4.00	0.85	10.1
1	4	10	1.44	3.61	-0.291	0.848	7.22	0.93	10.1
1	5	10	0.461	29.4	-8.25	20.1	39.5	1.02	10.0
3	6	30	10.0	0.97	2.60	0.015	2.62	2.75	29.0
3	9	30	7.18	1.64	1.75	0.11	4.00	2.54	30.5
3	12	30	4.31	3.61	-0.844	0.848	7.22	2.81	30.3
3	15	30	1.38	29.4	-24.7	20.1	39.5	3.04	29.8

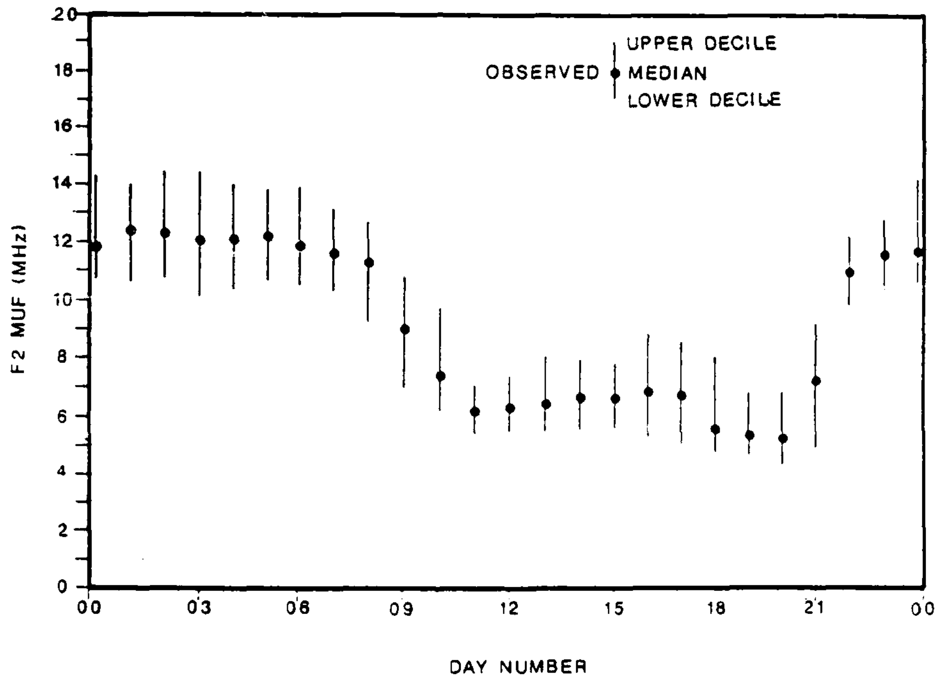

Figure 1 OTHR Maximum Usable Frequencies

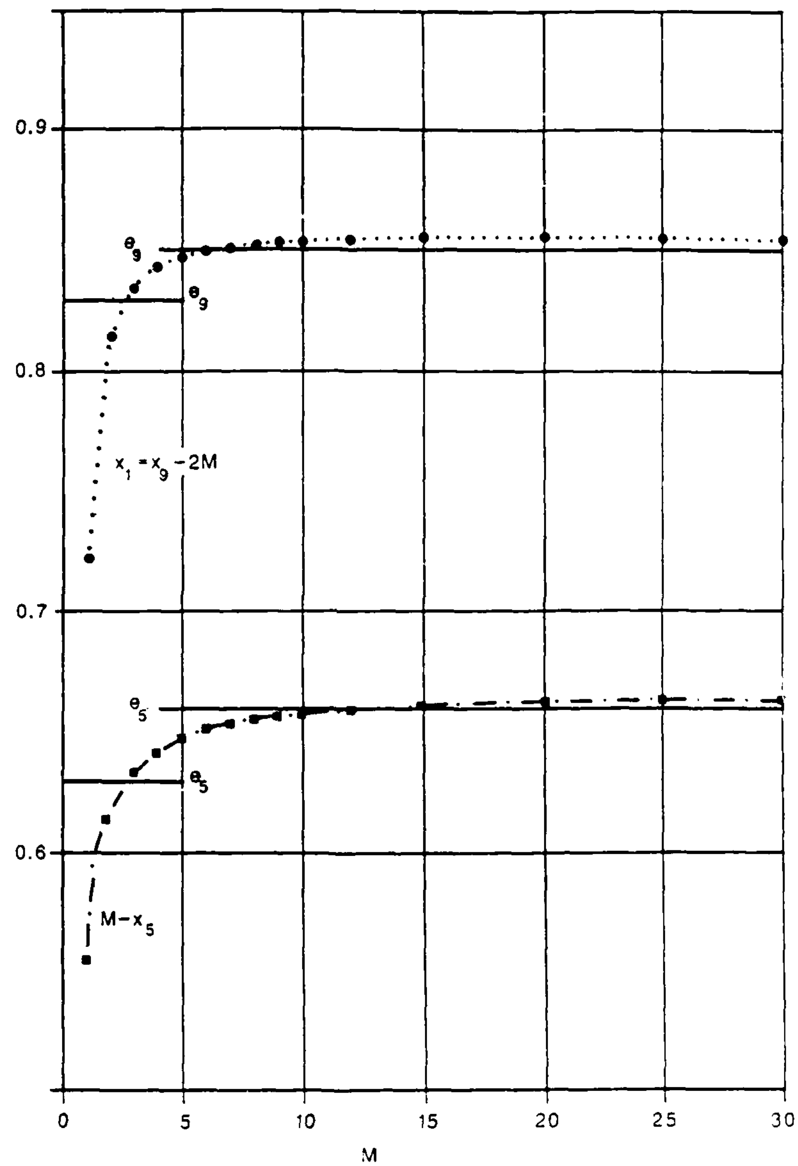

Figure 2 $M-x_5$ and $(x_1+x_9)-2M$

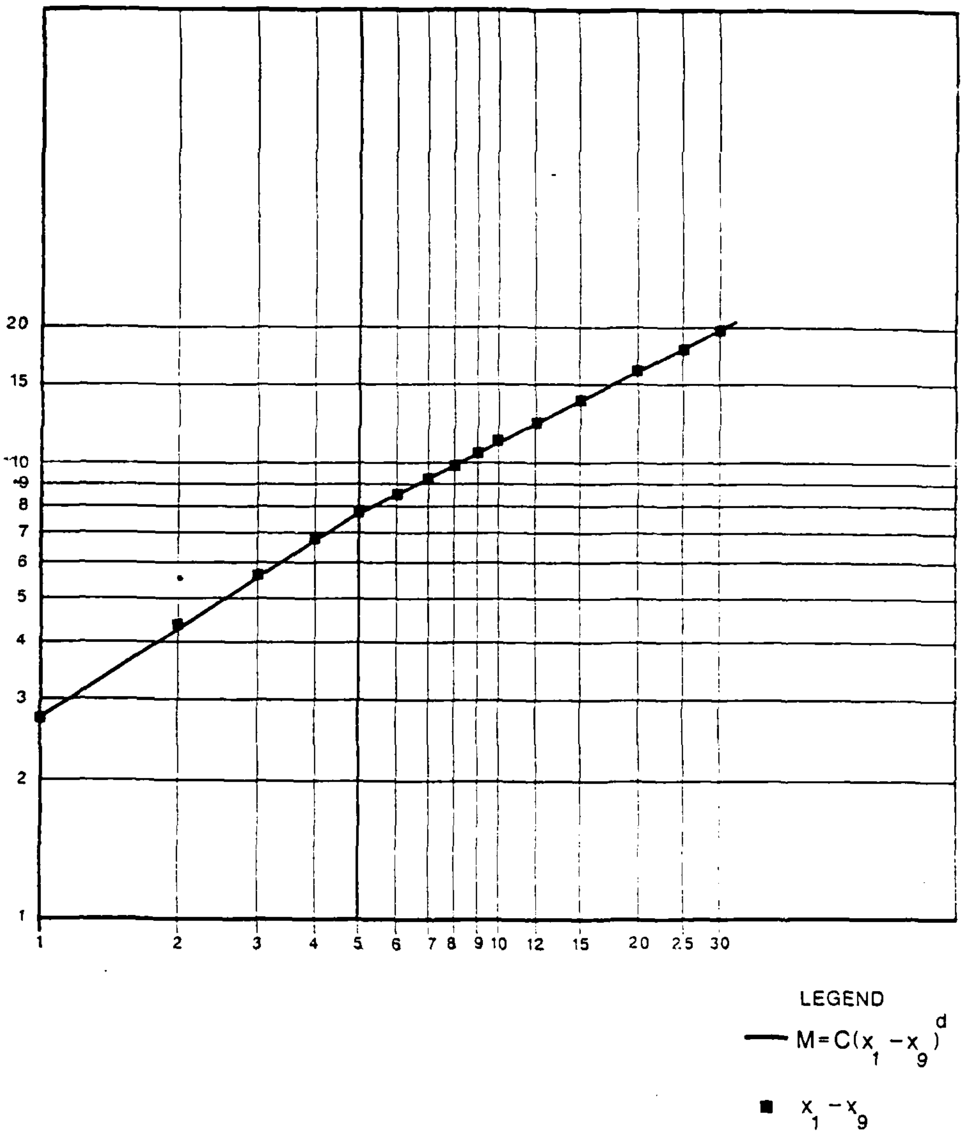

Figure 3 M vs $(x_1 - x_9)$ and $M = C\,(X_1 - X_9)^d$

DISTRIBUTION

	Copy No.
DEPARTMENT OF DEFENCE	
Central Office	
Deputy Secretary B	1
First Assistant Secretary, Force Development and Analysis Division	2
Headquarters Australian Defence Force	
Vice Chief of the Defence Force	3
Assistant Chief of the Defence Force (Policy)	4
Navy Office	
Navy Scientific Adviser	5
Director Operational Analysis - Navy	
Army Office	
Scientific Adviser - Army	6
Director Operational Analysis - Army	
Air Force Office	
Air Force Scientific Adviser	7
Director Operational Analysis - Air Force	
Defence Science and Technology Organisation	
Chief Defence Scientist	8
Assistant Chief Defence Scientist - Policy	9
Central Studies Branch	
Assistant Secretary, Central Studies	10
Principal Research Scientists, CSB (3)	11-13
Senior Service Representatives, CSB (3)	14-16
Author	17-26
Technical Reports Centre DCL-CP (7)	27-33

AL 149
Revised May 86

Department of Defence

DOCUMENT CONTROL DATA

Page Classification

1a. A R Number	1b. Establishment Number	2. Document Date	3. Task Number
AR-005-268	CSB-WP-RESEARCH11	JULY 1988	

4. Title	5. Security Classification (Place appropriate classification in box/s ie Secret(S), Confidential(C), Restricted (R), Unclassified (U))	6 No Pages	7 No Refs
CHI-SQUARED DISTRIBUTION PARAMETER APPROXIMATION	U document U title U abstract		

8. Author(s)	9. Downgrading/Delimiting Instructions
B.K. McMILLAN	-

10. Corporate Author and Address

Central Studies Branch,
Department of Defence,
CANBERRA, A.C.T. 2600
AUSTRALIA.

11 Office/Position responsible for ·

Sponsor CSB
Security CSB
Downgrading CSB
Approval CSB

12. Secondary Distribution *(Of this document)*

Approved for Public Release

Overseas enquiries outside stated limitations should be referred through ASDIS, Defence Information Services Branch, Department of Defence, Campbell Park, CANBERRA, ACT 2601

13a. This document may be announced in catalogues and awareness services available to...

No Limitations

13b. Citation for other purposes *(ie. casual announcement)* may be ☐ unrestricted or ☐ as for 13a

14. Descriptors	15. COSATI Group

16. Abstract

A method for approximating the mean of a chi-squared distribution and two linear transform constants from given values for the first, fifth and ninth deciles is presented. The origin of the work lies in OTHR modelling. Errors in the approximations are estimated and a range of example results tabulated.